

How to Achieve the Optimal DMT of Selective Fading MIMO Channels?

Lina Mroueh[†] and Jean-Claude Belfiore^{*}

[†] ISEP Paris, 75006 Paris, France

^{*} Télécom ParisTech, 75013 Paris, France

lina.mroueh@isep.fr and belfiore@enst.fr

Abstract

In this paper, we consider a particular class of selective fading channel corresponding to a channel that is selective either in time or in frequency. For this class of channel, we propose a systematic way to achieve the optimal DMT derived in Coronel and Bölcskei, *IEEE ISIT*, 2007 by extending the non-vanishing determinant (NVD) criterion to the selective channel case. A new code construction based on split NVD parallel codes is then proposed to satisfy the NVD parallel criterion. This result is of significant interest not only in its own right, but also because it settles a long-standing debate in the literature related to the optimal DMT of selective fading channels.

Index Terms

Diversity multiplexing tradeoff, selective channel, code construction, cyclic division algebra, non vanishing determinant (NVD) code.

I. INTRODUCTION AND MOTIVATIONS

In this paper, we consider the selective fading MIMO channel where a transmitter having n_t antennas wants to communicate with a receiver having n_r antennas. We assume that the communication occurs on a channel that exhibits memory either in time or in frequency. Our objective here is to construct reliable coding scheme for the high data rate communication in

Part of this work was performed while the first author was with the Communication Theory Group in ETH Zürich. Lina Mroueh is now at the department of signal, image and telecommunications in the Institut Supérieur d'Electronique de Paris ISEP, France and Jean-Claude Belfiore is with the department of communication and electronics in Telecom ParisTech, France. This paper was presented in part at the IEEE Information Theory Workshop (ITW) in Sept. 2010, Dublin, Ireland.

the high SNR regime when the channel is not known at the transmitter side. The performance criteria to evaluate the coding scheme that will be used in the following is the well-known diversity multiplexing tradeoff (DMT).

The diversity multiplexing tradeoff (DMT) proposed by Zheng and Tse in [1] is a powerful approach to characterizing the dual benefits in terms of diversity and spatial multiplexing in the high SNR regime. In order to achieve the optimal diversity multiplexing tradeoff for the flat fading MIMO channel, Belfiore *et al.* introduced the non-vanishing determinant criterion in [2]. Later, Elia *et al.* [3] proved that this criterion is a sufficient condition to achieve the optimal DMT using a full rate code.

While most of the above results address the case of flat fading channels, the general channel model of time-frequency selective channels has been considered by Coronel and Bölcskei in [4], [5] where the optimal DMT is derived. Moreover, a DMT optimal coding scheme based on a joint precoder and parallel codes construction, is proposed. As the block fading channel is a special case of the time-frequency selective channel, it is expected that the DMT expression in [5] matches with the corresponding result in [1]. This is, however, not the case¹ and has given rise to lots of debate in the literature *e.g.* [6]. A rigorous interpretation of this incoherence in results remains an open problem. The present paper settles the issue and shows that the DMT derived in [5] is, indeed, achievable.

Contributions: We consider a particular class of the general channel model considered in [4], [5] where the channel is selective either in time or in frequency. For this class of channels, we propose a systematic way to achieve the optimal DMT by extending the non-vanishing determinant criterion to the selective channel case. A new code construction based on split NVD parallel codes is then proposed to satisfy the NVD parallel criterion. Moreover, for the block fading channel, we provide an extension of the geometrical interpretation to show the achievability of the optimal DMT. This result is of significant interest not only in its own right, but also as it shows that the optimal DMT in [5] is achievable for all the classes of fading channels including the block fading channel.

Outline of the paper: The rest of the paper is organized as follows. In Section II, we define the selective fading channel model. We review in Section III some basic preliminaries and background materials that are essential to the development of this paper. Then, we derive in Section IV the limiting outage bound on the achievable DMT. We derive in Section V

¹The optimal DMT expression in [5] is larger than the one in [1].

the code design criterion required to achieve this optimal DMT for this class of selective channels and propose a new family of split NVD parallel codes to satisfy this code design criterion. Finally, Section VII concludes the paper.

Notation: The notation used in this paper is as follows. Boldface lower case letters \mathbf{v} denote vectors, boldface capital letters \mathbf{M} denote matrices. \mathbf{M}^\dagger denotes conjugate transposition. $.[^T]$ denote the transposition operator. $\|\mathbf{H}\|_F^2 = \text{Tr}\{\mathbf{H}\mathbf{H}^\dagger\}$ is the Frobenius norm of a matrix. $\text{Tr}\{\mathbf{A}\}$ refers to the trace of matrix \mathbf{A} . \mathbf{I}_N stands for the $N \times N$ identity matrix. $\text{diag}\{\mathbf{A}_n\}_{n=0}^{N-1}$ denotes the block diagonal matrix containing \mathbf{A}_n on its diagonal. $\text{vec}\mathbf{A} = [\mathbf{a}_1^{[T]} \dots \mathbf{a}_N^{[T]}]^{[T]}$, and \mathbf{a}_i is a column vector of matrix \mathbf{A} . The non zero eigenvalues of \mathbf{A} ordered in ascending order are denoted by $\lambda_i(\mathbf{A})$. \mathcal{CN} represents the complex Gaussian random variable. \mathbb{E}_X is the mathematical expectation w.r.t. to the random variable X . Equality in distribution between two random variables X and Y is represented by $X \sim Y$. Exponential equality is denoted by $f(x) \doteq x^b$, i.e. $\lim_{x \rightarrow \infty} \frac{\log f(x)}{\log x} = b$, and \gtrsim, \lesssim denote the exponential inequality. $|\mathcal{A}|$ denotes the cardinality of a set \mathcal{A} . Finally, $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of the matrices \mathbf{A} and \mathbf{B} .

II. CHANNEL AND SIGNAL MODEL

We consider the general case of selective fading channel which includes the case of time and frequency selective channel. In order to deal with such type of channels, techniques that decompose these channel into parallel sub-channels are generally used in literature [7]. The input-output relation for the class of channels considered in this paper is therefore given by

$$\mathbf{Y}_n^{[n_r \times T]} = \sqrt{\frac{\text{SNR}}{n_t}} \mathbf{H}_n^{[n_r \times n_t]} \mathbf{X}_n^{[n_t \times T]} + \mathbf{Z}_n^{[n_r \times T]}, \quad (1)$$

where $n = 0, 1, \dots, N-1$ represents the sub-channel n , the sub-channel $\mathbf{H}_n^{[n_r \times n_t]}$ is a $n_t \times n_r$ MIMO channel that remains constant during all the duration of the transmission T , \mathbf{X}_n represents the transmitted signal, and \mathbf{Z}_n denotes the additive i.i.d. $\mathcal{CN}(0, \mathbf{I})$ noise. The channels \mathbf{H}_n are correlated across the sub-channels $n = 0 \dots N-1$ according to,

$$\mathbf{H} = [\mathbf{H}_0 \dots \mathbf{H}_{N-1}] = \mathbf{H}_w (\mathbf{R}_{\mathbb{H}}^{1/2} \otimes \mathbf{I}_{n_t}), \quad (2)$$

where $\mathbf{R}_{\mathbb{H}}$ is the $N \times N$ correlation between the scalar sub-channels characterized by its rank equal to $\rho \leq N$, \mathbf{H}_w is an $n_r \times Nn_t$ matrix with i.i.d. $\mathcal{CN}(0, 1)$ entries. The transmitted signal satisfies the following power constraint,

$$\sum_{i=0}^{N-1} \mathbb{E} [\|\mathbf{X}_i\|_F^2] \leq TN. \quad (3)$$

Throughout this paper, we set $m = \min(n_t, n_r)$ and $M = \max(n_t, n_r)$.

The input-output relation considered in (1) models the case when the channel is selective either in time or in frequency. For the frequency selective channel, the MIMO OFDM system decomposes the channel into N parallel subcarrier, where N represents the total number of subcarriers and n stands for the frequency. The sub-channel remains constant over each subcarrier and the correlation matrix $\mathbf{R}_{\mathbb{H}}$ is a circulant matrix having a rank equal to L which is nothing but the number of channel taps or the *memory* of the selective channel.

For the time selective case (or the block fading channel), the channel remains constant during a block n of T time slots and changes in a statistically independent manner across blocks. For this case, N represents the total number of blocks and $\mathbf{R}_{\mathbb{H}} = \mathbf{I}_N$ with full rank N .

III. PRELIMINARIES AND BACKGROUND

In this section, we start by recalling some basic preliminaries on the optimal diversity multiplexing tradeoff (DMT) of the code in Subsection III-A and on the limiting outage bound of the selective fading channel in Subsection III-B. Then, we briefly review prior results from literature that motivate our contribution.

A. Diversity multiplexing tradeoff (DMT)

Let $\mathcal{X}_p(\text{SNR})$ be a family of coding schemes operating at a given SNR, and let $R(\text{SNR})$ denote the rate transmitted *per sub-channel*, such that,

$$R(\text{SNR}) = r \log \text{SNR},$$

where r is the multiplexing gain *per sub-channel*.

The diversity multiplexing tradeoff (DMT) of the coding scheme $\mathcal{X}_p(\text{SNR})$ is defined as the SNR exponent of the error probability $P_{e,\mathcal{X}_p}(r, \text{SNR})$ using maximum likelihood-decoding such that

$$d(r) = - \lim_{\text{SNR} \rightarrow \infty} \frac{\log P_{e,\mathcal{X}_p}(r, \text{SNR})}{\log \text{SNR}}.$$

For a given multiplexing gain r , the optimal DMT is the largest DMT supported by any coding scheme, and is denoted by $d^*(r)$.

B. DMT outage bound

The outage probability of a selective fading channel when the target rate R scales as $r \log \text{SNR}$ is defined as,

$$P_{\text{out}}(r) \triangleq \mathbb{P} \left\{ \log \det \left(\mathbf{I}_N + \frac{\text{SNR}}{n_t} \mathbf{H} \mathbf{H}^\dagger \right) < Nr \log \text{SNR} \right\},$$

where $\mathbf{H} = \text{diag}\{\mathbf{H}_n\}_{n=0}^{N-1}$ is the block diagonal channel matrix.

The optimal DMT of the selective fading MIMO channel has been derived in [4] and [5]. For this general case, Coronel and Bölcskei showed that the outage probability is bounded as,

$$P_{e,\mathcal{X}_p}(r) \geq P_{\text{out}}(r) \geq P_J(r) \doteq \text{SNR}^{-d_J(r)} \quad (4)$$

where,

$$d_J(r) = (\rho M - r)(m - r). \quad (5)$$

and $m = \min(n_t, n_r)$ and $M = \max(n_t, n_r)$. Note that the first inequality in (4) is a consequence of the Fano inequality [1], and the second inequality is a consequence of the Jensen inequality as shown in [4]. Moreover, a coding scheme that achieves the bound called "Jensen bound" in the terminology of [5] has been proposed in [4] and [5]. It follows therefore from [5] that the optimal DMT is equal to,

$$d(r) = d_J(r) = (\rho M - r)(m - r).$$

C. Previous work and motivations

The block fading channel is a particular case of the selective fading channel model considered in (1) with covariance matrix $\mathbf{R}_{\mathbb{H}} = \mathbf{I}_N$. The optimal DMT expression is therefore $d^*(r) = (NM - r)(m - r)$, which is the DMT expression of the general channel model considered in [4], [5] applied to this particular channel setting. Obviously, this result does not match with the corresponding result in [1], *i.e.*, $d_l(r) = N(M - r)(m - r) \leq d^*(r)$, $\forall r$. This incoherence in results has been subject to lots of debate in literature *e.g.* [6] and motivates our contribution. The authors of [6] base their arguments on a non-accurate outage probability derivation ($P_{\text{out},l}(r) \doteq \text{SNR}^{-d_l(r)}$) to claim that the DMT of the block fading channel cannot exceed $d_l(r) \leq d^*(r)$. In order to settle this issue, we show in this paper that the DMT in [1] is not a limiting outage bound as claimed in [6], and that the DMT in [5] is achievable using codes derived from cyclic division algebra (CDA).

IV. OUTAGE BOUND ON THE DMT OF SELECTIVE FADING CHANNEL

Unlike the flat fading channel, the analytical outage probability for the selective fading channel cannot be easily derived using the eigenvalues distribution. For the case of correlated parallel sub-channels, Coronel and Bölcskei in [4] generalize the geometrical interpretation in [1] to the selective fading case. For the particular case of the statistically independent parallel sub-channels which is the block fading channel, the analytical outage probability should be carefully performed to take into account the impact of coding across the blocks, which cannot be easily seen using the block diagonal structure of the matrix. For this, an equivalent expression of the outage probability is first derived. Then, we provide here an outage derivation based on the geometrical argument previously used for the flat fading channel in [1] and for the selective fading case in [4].

A. Outage bound of the block fading channel

For the block fading channel, the outage probability is,

$$P_{\text{out}}(r) \triangleq \mathbb{P} \left\{ \log \det \left(\mathbf{I} + \frac{\text{SNR}}{n_t} \mathcal{H} \mathcal{H}^\dagger \right) < Nr \log \text{SNR} \right\},$$

where $\mathcal{H} = \text{diag}\{\mathbf{H}_n\}_{n=0}^{N-1}$ is the block diagonal channel matrix.

1) *Equivalent outage expression:* In order to generalize the geometrical interpretation in [4] to the block fading channel, we start first by finding in Lemma 1 an equivalent expression of the outage probability.

Lemma 1: For the block fading channel, the outage probability is equivalent to,

$$P_{\text{out}}(r) = \mathbb{P} \left\{ \log \det \left(\mathbf{I} + \frac{\text{SNR}}{Nn_t} \mathbf{C}_H \mathbf{C}_H^\dagger \right) < Nr \log \text{SNR} \right\}, \quad (6)$$

where

$$\mathbf{C}_H = \begin{bmatrix} \mathbf{H}_{w,0} & \mathbf{H}_{w,1} & \dots & \mathbf{H}_{w,N-1} \\ & & \vdots & \\ \mathbf{H}_{w,1} & \mathbf{H}_{w,2} & \dots & \mathbf{H}_{w,0} \end{bmatrix} \quad (7)$$

and $\mathbf{H}_{w,i}$, $i = 0 \dots N-1$ are Gaussian matrices with i.i.d. entries.

Before going to the rigorous proof, we note here that the main intuition behind this lemma is the fact that the block fading channel can be considered as a selective fading channel with a channel memory of N blocks. This is so far the case as the covariance matrix is equal to identity, which is a full rank matrix with rank equal to N .

Proof: To prove this lemma, we consider \mathbf{h}_{ij} the $N \times 1$ Gaussian vector $\sim \mathcal{CN}(0, \mathbf{I}_N)$ containing the N independent channel realisations between transmit antenna j and receive antenna i . It is well-known that the Gaussian vector \mathbf{h}_{ij} is identically distributed as $\mathbf{F}\mathbf{h}_{\omega,ij}$ for any unitary matrix \mathbf{F} , i.e., $\mathbf{h}_{ij} \sim \mathbf{F}\mathbf{h}_{\omega,ij}, \forall i, j$.

In the following, we specify our result to the case where F is a $N \times N$ Fast Fourier Transform (FFT) matrix. This means that each channel realisation is identically distributed as,

$$h_{ij}^{[n]} \sim \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} h_{ij,w}^{[l]} e^{-j2\pi \frac{ln}{N}}, \quad n = 0 \dots N-1.$$

The block diagonal matrix \mathcal{H} is therefore identically distributed as \mathbf{D}_H , i.e., $\mathcal{H} \sim \mathbf{D}_H$, where,

$$\mathbf{D}_H = \frac{1}{\sqrt{N}} \begin{bmatrix} \sum_{l=0}^{N-1} \mathbf{H}_{w,l} \omega_l^0 & & \\ & \ddots & \\ & & \sum_{l=0}^{N-1} \mathbf{H}_{w,l} \omega_l^{N-1} \end{bmatrix}, \quad (8)$$

with $\omega_l = e^{-j\frac{2\pi l}{N}}$ and $\mathbf{H}_{\omega,l} = (h_{ij,\omega}^{[l]})_{1 \leq i \leq n_r, 1 \leq j \leq n_t}$.

Consequently, the mutual information is identically distributed as,

$$I(\mathbf{x}, \mathbf{y} | \mathbf{H}) \sim \log \det \left(\mathbf{I} + \frac{\text{SNR}}{N n_t} \mathbf{D}_H \mathbf{D}_H^\dagger \right) = I_D(\text{SNR}).$$

By using an FFT precoder and an FFT equalizer as in an OFDM system to transmit over the channel \mathbf{D}_H in (8), the matrix $\mathbf{D}_H \mathbf{D}_H^\dagger$ can be made unitarily equivalent to $\mathbf{C}_H \mathbf{C}_H^\dagger$, where

$$\mathbf{C}_H = \begin{bmatrix} \mathbf{H}_{w,0} & \mathbf{H}_{w,1} & \dots & \mathbf{H}_{w,N-1} \\ \mathbf{H}_{w,N-1} & \mathbf{H}_{w,0} & \dots & \mathbf{H}_{w,N-2} \\ & & \ddots & \\ \mathbf{H}_{w,1} & \mathbf{H}_{w,2} & \dots & \mathbf{H}_{w,0} \end{bmatrix}.$$

Thus, the corresponding mutual information $I_D(\text{SNR})$ can be written as,

$$I_D(\text{SNR}) = \log \det \left(\mathbf{I} + \frac{\text{SNR}}{N n_t} \mathbf{C}_H \mathbf{C}_H^\dagger \right) \sim I(\mathbf{x}, \mathbf{y} | \mathbf{H}).$$

It follows therefore that the outage probability is such that,

$$\mathbf{P}_{\text{out}}(r) = \mathbb{P} \left\{ \log \det \left(\mathbf{I} + \frac{\text{SNR}}{N n_t} \mathbf{C}_H \mathbf{C}_H^\dagger \right) < N r \log \text{SNR} \right\}.$$

■

2) *Geometrical interpretation:* Following the geometrical interpretation of the flat fading channel in [1], the typical outage event occurs when the channel matrix \mathbf{C}_H is close to the manifold of all matrices with rank Nr denoted by \mathcal{R}_{Nr} , such that,

$$\mathcal{R}_{Nr} = \{\mathbf{C}_H : \text{rank}\{\mathbf{C}_H\} = Nr\}.$$

By following the same reasoning as in [1], this requires that the $d(r)$ components of \mathbf{C}_H orthogonal to \mathcal{R}_{Nr} to be collapsed, *i.e.*, be on the order of SNR^{-1} . The probability of this event is $P_{\text{out}}(r) \doteq \text{SNR}^{-d(r)}$. The number of these components is given by

$$d(r) = NMm - \dim(\mathcal{R}_{Nr}),$$

where $\dim(\mathcal{R}_{Nr})$ is the sufficient *minimal* number of parameters required to specify matrix \mathbf{C}_H with rank Nr .

3) *Dimensionality of \mathcal{R}_{Nr} :* We first note that due to the structure of \mathbf{C}_H in (7), the number of parameters required to characterize a matrix \mathbf{C}_H in \mathcal{R}_{Nr} is equal to the number of parameters required to specify an $m \times NM$ matrix ($m = \min(n_t, n_r)$ and $M = \max(n_t, n_r)$) with rank r that contains the n_t first columns if $n_t \leq n_r$, and the n_r first rows if $n_r \leq n_t$ as shown in Figures 1(a) and 1(b).

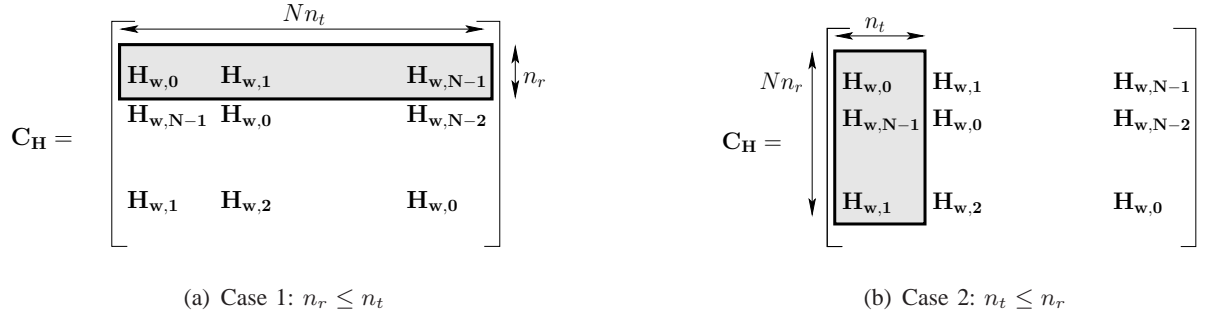


Fig. 1. It is sufficient to specify a $m \times NM$ matrix with rank r , with $m = \min(n_t, n_r)$, $M = \max(n_t, n_r)$ to characterize a matrix \mathbf{C}_H with rank Nr .

Characterizing a matrix \mathbf{C}_H with rank Nr reduces therefore to the problem of characterizing a matrix of dimension $m \times NM$ with rank r that requires only $NMr + (m - r)r$, *i.e.*,

$$\dim(\mathcal{R}_{Nr}) = NMr + (m - r)r,$$

where NMr is the number of independent parameters needed to identify r independent vectors and $(m - r)r$ parameters are needed to identify the linear dependent vectors as a function of the r independent vectors. It can be easily verified here that the NMr free

i.i.d. Gaussian parameters that identify the r linear independent vectors generate a block circulant matrix with rank Nr with a probability equal to one.

It can be deduced that the optimal DMT for the class of block fading channel is,

$$d_{out}(r) = NMm - \dim(\mathcal{R}_{Nr}) = (NM - r)(m - r).$$

B. Comments on related work's derivation

It turns out from the geometrical interpretation that the outage event is reduced to the probability that the $m \times NM$ Jensen channel, denoted by \mathcal{H}_w in the rest of the paper, is in outage, which is the Jensen outage event in the Coronel and Bölcskei terminology [5]. This means that the outage event is reduced to,

$$\mathcal{O}(\text{SNR}) = \{\mathcal{H}_w \in \mathbb{C}^{m \times NM} \text{ is in outage}\}$$

Note that the straightforward generalization of the flat fading outage results to the block diagonal matrix in (6) as in [1] and [6] does not take into account the impact of the coding among the channel blocks in the analytical outage derivation and does not lead to an accurate outage probability expression. In the following, we show how this optimal DMT can be achieved using a code derived from cyclic division algebra (CDA).

V. DMT ACHIEVABILITY: SPLIT NVD PARALLEL CODES FOR SELECTIVE FADING CHANNEL

In this section, we propose a new family of split NVD parallel codes to achieve the optimal DMT of $(\rho M - r)(m - r)$. We start first by deriving in Subsection V-A a sufficient condition on the code to achieve the optimal DMT for this class of channel. The new family of proposed codes is based on the previously known NVD parallel codes family which we will briefly review in Subsections V-B and V-C. Finally, the code construction and the optimality of the split NVD parallel code is addressed in Subection V-E.

A. Optimal code design criterion

Unlike the case of time-frequency selective channel in [4], we show here that when the channel is selective either in time or in frequency, there is no need to construct an additional precoder adapted to the channel statistics in order to achieve the optimal DMT. The optimal code design criterion required to achieve the optimal DMT is summarized in the following theorem.

Theorem 1 (Sufficient condition for DMT achievability): A coding scheme $\mathbf{X} \in \mathcal{X}_p(\text{SNR})$ achieves the optimal DMT $(\rho M - r)(m - r)$, if for any two different codewords $\mathbf{X}, \hat{\mathbf{X}} \in \mathcal{X}_p(\text{SNR})$, the eigenvalues of the block diagonal matrix $\mathbf{D}\mathbf{D}^\dagger$, where $\mathbf{D} = \text{diag}\{(\mathbf{X}_n - \hat{\mathbf{X}}_n)\}_{n=0}^{N-1}$ satisfy

$$\min_{\mathbf{X}, \hat{\mathbf{X}} \in \mathcal{X}_p(\text{SNR})} \prod_{i=1}^m \lambda_i(\mathbf{D}\mathbf{D}^\dagger) \geq \frac{1}{2^{R(\text{SNR}) + o(\text{SNR})}}. \quad (9)$$

Proof: The proof of this theorem uses the same steps as the proof of [Theorem 1 in [4]] and is detailed in Appendix A. ■

B. NVD parallel scheme

Let $\mathbf{X} = \text{diag}\{\mathbf{X}_n\}_{n=0}^{N-1} \in \mathcal{X}_p(\text{SNR})$ be the block diagonal matrix containing the transmitted codeword \mathbf{X}_i in (1), and constructed such that $\mathbf{X} = \theta \mathbf{\Xi}$, where θ is a scaling factor that depends on the structure of the code, and chosen to ensure the power constraint in (3). The block diagonal matrix $\mathbf{\Xi} = \text{diag}\{\tilde{\mathbf{\Xi}}_i\}_{i=0}^{N-1}$ is an NVD parallel code denoted by $\mathcal{C}(\text{SNR})$, and defined as follows:

Definition 1 (NVD parallel scheme): Let $\mathcal{A}(\text{SNR})$ be an alphabet² that is salably dense, such that

$$\forall s \in \mathcal{A}(\text{SNR}) \Rightarrow |s|^2 \leq |\mathcal{A}(\text{SNR})|.$$

Then, $\mathcal{C}(\text{SNR})$ is called NVD parallel code if,

- 1) Each entry of $\mathbf{\Xi}$ is a linear combination of symbols carved from $\mathcal{A}(\text{SNR})$.
- 2) The total number of transmitted symbols carved from $\mathcal{A}(\text{SNR})$ is equal to TNn_t .
- 3) For any pair of different codewords $\mathbf{\Xi}$ and $\hat{\mathbf{\Xi}} \in \mathcal{C}(\text{SNR})$, the NVD property is satisfied

$$\det((\mathbf{\Xi} - \hat{\mathbf{\Xi}})(\mathbf{\Xi} - \hat{\mathbf{\Xi}})^\dagger) \geq \kappa > 0, \quad (10)$$

with κ is a constant independent of SNR.

²We assume here without restriction that the signal constellation is a QAM constellation, i.e., $\mathcal{A}(\text{SNR}) = \mathcal{A}_{\text{QAM}}(\text{SNR})$. This can be also extended to the case of HEX constellations.

C. Cyclic division algebra (CDA) code structure

We recall here the most relevant concepts of the construction of the codeword matrix $\Xi = \text{diag}\{\tilde{\Xi}_i\}_{i=0}^{N-1}$ based on cyclic division algebra. We refer the reader to [8], [9] for more details on the NVD parallel code construction. In the following, we consider,

- The field \mathbb{F} as a Galois extension of degree N over $\mathbb{Q}(i)$, and that have τ as generator, such that

$$\text{Gal}(\mathbb{F}/\mathbb{Q}(i)) = \{\tau_0, \dots, \tau_{N-1}\}.$$

- The field \mathbb{K} is a cyclic extension of degree n_t over \mathbb{F} , and that have σ as generator, such that

$$\text{Gal}(\mathbb{K}/\mathbb{F}) = \{\sigma^0, \dots, \sigma^{n_t-1}\}.$$

The code Ξ is constructed by setting $\tilde{\Xi}_i = \tau_i(\tilde{\Xi})$, *i.e.*,

$$\Xi = \begin{bmatrix} \tilde{\Xi} & & & \\ & \tau_1(\tilde{\Xi}) & & \\ & & \dots & \\ & & & \tau_{N-1}(\tilde{\Xi}) \end{bmatrix} \quad (11)$$

where $\tilde{\Xi}$ belongs to the cyclic division algebra $\mathcal{C} = (\mathbb{K}/\mathbb{F}, \sigma, \gamma)$, and $\gamma \in \mathbb{F}$ chosen such that $\gamma, \gamma^2, \dots, \gamma^{n_t-1}$ are not norms of an element of \mathbb{K} . The matrix $\tilde{\Xi}$ is defined such that

$$\tilde{\Xi} = \begin{pmatrix} x_0 & x_1 & \dots & x_{n_t-1} \\ \gamma\sigma(x_{n_t-1}) & \sigma(x_0) & \dots & \sigma(x_{n_t-2}) \\ \vdots & & & \vdots \\ \gamma\sigma^{n_t-1}(x_1) & \gamma\sigma^{n_t-1}(x_2) & \dots & \sigma^{n_t-1}(x_0) \end{pmatrix},$$

where, $x_i = \sum_{j=1}^{Nn_t} s_{i,j}\omega_j$, $s_{i,j} \in \mathcal{A}(\text{SNR})$ and $\omega_j \in \mathbb{K}$. For the NVD parallel code, the determinant is such that,

$$\begin{aligned} \det(\text{diag}\{\tilde{\Xi}_i\}_{i=1}^N) &= \prod_k \tau_k(\det(\tilde{\Xi})) \\ &= N_{\mathbb{F}/\mathbb{Q}(i)}(\det(\tilde{\Xi}_i)) \in \mathbb{Z}[i], \end{aligned}$$

and which is equal to zero if and only if all x_i are zeros. It follows that for $\Xi \neq 0$,

$$|\det(\Xi)|^2 \geq \text{SNR}^0.$$

We finally recall that the NVD parallel codes preserve the mutual information as,

$$\text{vec}\left(\begin{bmatrix} \tilde{\Xi}^{[T]} & \dots & \tau_{N-1}(\tilde{\Xi})^{[T]} \end{bmatrix}^{[T]}\right) = \Phi \mathbf{s}$$

where Φ is an orthogonal matrix, such that $\Phi\Phi^\dagger = \mathbf{I}_{Nn_t}$. It follows therefore that the mutual information between the vectorized input vectors $\tilde{\mathbf{x}} = \text{vec}(\begin{bmatrix} \mathbf{X}_0^{[T]} \dots \mathbf{X}_{N-1}^{[T]} \end{bmatrix})$ and the vectorized output $\tilde{\mathbf{y}} = \text{vec}(\begin{bmatrix} \mathbf{Y}_0^{[T]} \dots \mathbf{Y}_{N-1}^{[T]} \end{bmatrix})$ is,

$$I(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}|\mathbf{H}) = \log \det \left(\mathbf{I}_N + \frac{\text{SNR}}{n_t} \mathcal{H} \mathcal{H}^\dagger \right),$$

where $\mathcal{H} = \text{diag}\{\mathbf{H}_n\}_{n=0}^{N-1}$ is the block diagonal channel matrix.

D. Choice of θ for NVD parallel codes

Following the same reasoning in [3] and [6], the scaling factor θ that insures the power constraint in (3) is such that,

$$\theta^2 \sum_{i=0}^{N-1} \mathbb{E}[\|\tilde{\mathbf{\Xi}}_i\|_{\text{F}}^2] \leq TN.$$

Due the linearity of this code and to the use of unit transformation, each entry of $x \in \Xi$ is such that,

$$\begin{aligned} \mathbb{E}[|x|^2] &= \mathbb{E}[|s|^2], \quad s \in \mathcal{A}_{\text{QAM}}(\text{SNR}), \\ &= \frac{2(|\mathcal{A}(\text{SNR})| - 1)}{3}. \end{aligned}$$

This implies that,

$$\begin{aligned} \sum_{i=0}^{N-1} \mathbb{E}[\|\tilde{\mathbf{\Xi}}_i\|_{\text{F}}^2] &= TN \mathbb{E}[|x|^2], \\ &\doteq TN|\mathcal{A}(\text{SNR})|. \end{aligned}$$

The scaling factor θ that ensures the power constraint is therefore,

$$\theta^2 \doteq |\mathcal{A}(\text{SNR})|^{-1}. \quad (12)$$

Using the NVD parallel criterion in (10) and the value of θ^2 in (12), the eigenvalues of the block diagonal matrix $\mathbf{D} = \mathbf{X} - \hat{\mathbf{X}} = \theta(\Xi - \hat{\Xi})$ for any different codewords $\mathbf{X}, \hat{\mathbf{X}}$, are such that,

$$\prod_{i=1}^{Nn_t} \lambda_i(\mathbf{D}\mathbf{D}^\dagger) = \frac{|\det(\Xi - \hat{\Xi})|^2}{|\mathcal{A}(\text{SNR})|^{Nn_t}} \geq \frac{1}{|\mathcal{A}(\text{SNR})|^{Nn_t}}.$$

Due to the power constraint in (3), these eigenvalues necessarily satisfy $\lambda_i(\mathbf{D}\mathbf{D}^\dagger) \leq 1$. Then, the NVD parallel criterion is equivalent to,

$$\min_{\mathbf{X}, \hat{\mathbf{X}} \in \mathcal{X}_p(\text{SNR})} \prod_{i=1}^m \lambda_i(\mathbf{D}\mathbf{D}^\dagger) \geq \frac{1}{|\mathcal{A}(\text{SNR})|^{Nn_t}}. \quad (13)$$

It can be easily verified that the NVD parallel criteria of the NVD parallel code depends critically on the size of the constellation. The natural question that comes here is: What is the optimal size of constellation that guarantees to transmit a rate $R(\text{SNR})$ over each sub-channel and that meets the sufficient condition of DMT achievability in (9).

E. Split NVD parallel codes and optimality

The NVD parallel codes as put straightforwardly by Lu in [9] and Yang *et al.* in [8] are sub-optimal, as the DMT achieved by these codes is only $\rho(n_t - r)(n_r - r) < (\rho M - r)(m - r)$. The main idea of the new split code construction is to design a coding scheme that guarantees to transmit a rate of $R(\text{SNR})$ using a total power of SNR over each sub-channel and to satisfy the NVD parallel criterion in Theorem 1. The two possible ways of splitting the data over the parallel channels are detailed in Subsections V-E1 and V-E2.

1) *Block diagonal NVD parallel code:* The first way of splitting the data over the parallel channels has been previously studied in [9] and is depicted in Figure 2.

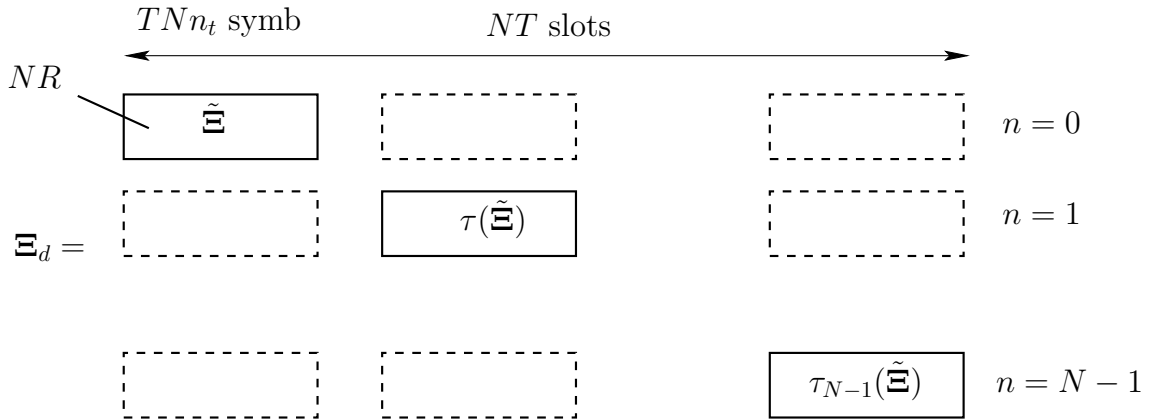


Fig. 2. Coding across time and frequency: The total rate is transmitted only during T slots. Each entry of $\tau_i(\tilde{\Xi})$ is a linear combination of symbols carved from $\mathcal{A}_d(\text{SNR})$ where $|\mathcal{A}_d(\text{SNR})| = \text{SNR}^{\frac{r}{n_t}}$. In this case, $\mathbf{X}_{e,d} = \theta_d \Xi_d$.

In this case, the total rate NR is transmitted during only T slots over each sub-channel. It can be easily verified that for this scheme the outage event is such that,

$$\mathcal{O}_1(r, \text{SNR}) = \{I_1(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}|\mathbf{H}) < Nr \log \text{SNR}\},$$

where,

$$I_1(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}|\mathbf{H}) = \log \det \left(\mathbf{I}_N + \frac{\text{SNR}}{n_t} \mathbf{H} \mathbf{H}^\dagger \right).$$

Each block $\tau_i(\tilde{\Xi})$ contains TNn_t symbols carved from a signal constellation $\mathcal{A}_d(\text{SNR})$. In order to maintain a rate of $R(\text{SNR})$ over each sub-channel, the size of the constellation $|\mathcal{A}_d(\text{SNR})|$ should be chosen such that,

$$R(\text{SNR}) = r \log \text{SNR} = \frac{1}{NT} \log |\mathcal{A}_d(\text{SNR})|^{n_t TN}.$$

i.e., $|\mathcal{A}_d(\text{SNR})| = \text{SNR}^{\frac{r}{n_t}}$. It can easily be verified that for this choice of signal constellation size, the NVD parallel criterion in (13) is,

$$\min_{\mathbf{X}, \hat{\mathbf{X}} \in \mathcal{X}_p(\text{SNR})} \prod_{i=1}^m \lambda_i(\mathbf{D}\mathbf{D}^\dagger) \geq \frac{1}{2^{NR(\text{SNR})+o(\text{SNR})}}.$$

Obviously, the sufficient condition in Theorem 1 is not satisfied in this case. The achievable DMT by this transmission scheme is only $\rho(n_t - r)(n_r - r)$ as shown in [9], and it is therefore sub-optimal.

2) *Split NVD parallel code*: The second way we propose to split the data that guarantees to transmit a rate of $R(\text{SNR})$ using a total power of SNR over each sub-channel is shown in Figure 3. In this case, the total rate is split equally among all the NT slots. Each block Ξ_i transmits TNn_t symbols carved from a signal constellation $\mathcal{A}_s(\text{SNR})$. The same TNn_t symbols are transmitted over blocks $\Xi_i \dots \tau_{N-1}(\Xi_i)$ but encoded differently. However, different symbols are transmitted over two different blocks Ξ_i and Ξ_j .

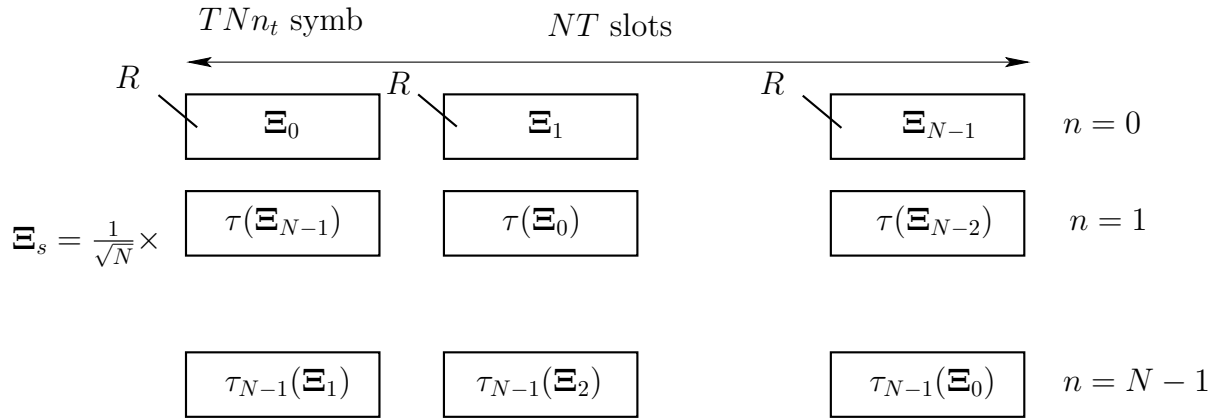


Fig. 3. Coding across time and frequency: The total rate is split across the NT slots. Each entry of $\tau_i(\Xi_i)$ is a linear combination of symbols carved from $\mathcal{A}_s(\text{SNR})$ where $|\mathcal{A}_s(\text{SNR})| = \text{SNR}^{\frac{r}{Nn_t}}$. In this case, $\mathbf{X}_{e,s} = \theta_s \Xi_s$.

For this transmission scheme, the outage event occurs when at least one of the NVD parallel code scheme with rate $R(\text{SNR}) = r \log \text{SNR}$ is in outage, meaning that,

$$\mathcal{O}_2(r, \text{SNR}) = \bigcup_{s=0}^{N-1} \mathcal{O}_s(r, \text{SNR}),$$

where,

$$\mathcal{O}_s(r, \text{SNR}) = \left\{ \frac{1}{N} I_2(\tilde{\mathbf{x}}, \tilde{\mathbf{y}} | \mathbf{H}) < r \log \text{SNR} \right\}, \quad \forall s,$$

and,

$$I_2(\tilde{\mathbf{x}}, \tilde{\mathbf{y}} | \mathbf{H}) = \log \det \left(\mathbf{I}_N + \frac{\text{SNR}}{N n_t} \mathbf{H} \mathbf{H}^\dagger \right).$$

Note that the normalization factor $1/N$ in the first side of the inequality in the outage event $\mathcal{O}_s(r, \text{SNR})$ traduces the fact that N blocks are needed to decode the information of each NVD parallel code with rate $R(\text{SNR})$.

Using the union bound and the inclusion bound ($\mathcal{O}_s \subseteq \mathcal{O}_2$), the outage probability can be bounded as,

$$\text{P}(\mathcal{O}_s) \leq \text{P}(\mathcal{O}_2) \leq \sum_{i=0}^{N-1} \text{P}(\mathcal{O}_s) \quad (14)$$

Assuming that $\text{P}(\mathcal{O}_s)$ scales as $\text{SNR}^{-d_s(r)}$, it follows from (14) that at high SNR,

$$\text{P}(\mathcal{O}_2) \doteq \text{SNR}^{-d_s(r)} \doteq \text{P}(\mathcal{O}_s) \doteq \text{P}(\mathcal{O}_1),$$

This implies that this scheme is equivalent in term of outage to the first scheme.

In order to maintain the rate of $R(\text{SNR})$ over each sub-channel, the signal constellation $\mathcal{A}_s(\text{SNR})$ should be chosen such that,

$$R(\text{SNR}) = r \log \text{SNR} = \frac{1}{T} \log |\mathcal{A}_s(\text{SNR})|^{n_t T N}.$$

The size of the signal constellation for the split NVD parallel scheme is therefore reduced compared to the block diagonal case, and

$$|\mathcal{A}_s(\text{SNR})| = \text{SNR}^{\frac{r}{N n_t}} = |\mathcal{A}_d(\text{SNR})|^{\frac{1}{N}}.$$

Due to the block diagonal channel matrix structure, it can be deduced that the split NVD parallel code is equivalent to a concatenation of N independent parallel NVD codes, where the symbols of each NVD parallel code are carved from a constellation $\mathcal{A}_s(\text{SNR})$ with size $\text{SNR}^{\frac{r}{N n_t}}$. The system is in error if at least one of the NVD parallel codes is in error, *i.e.*,

$$\varepsilon(r, \text{SNR}) = \bigcup_{i=0}^{N-1} \varepsilon_i(r, \text{SNR}),$$

where $\varepsilon(r, \text{SNR})$ represents the event that the system is in error and $\varepsilon_i(r, \text{SNR})$ denotes the event that the i^{th} NVD parallel code formed by the blocks $\Xi_i \dots \tau_{N-1}(\Xi_i)$ is in error. For each NVD parallel code with symbols carved from $\mathcal{A}_s(\text{SNR})$, it can be easily verified by

replacing the cardinality of $\mathcal{A}_s(\text{SNR})$ in (13) that the NVD parallel criterion in Theorem 1 is satisfied, *i.e.*,

$$\min_{\mathbf{X}, \hat{\mathbf{X}} \in \mathcal{X}_p(\text{SNR})} \prod_{i=1}^m \lambda_i(\mathbf{D}\mathbf{D}^\dagger) \stackrel{\cdot}{\geq} \frac{1}{2^{R(\text{SNR})+o(\text{SNR})}}.$$

It follows from Theorem 1 that,

$$\mathbf{P}(\varepsilon_i) \doteq \text{SNR}^{-d_i(r)},$$

where $d_i(r) = (\rho M - r)(m - r)$, $\forall i$.

Using the inclusion and the union bound as for the outage analysis in (14), it follows that,

$$\mathbf{P}_e(r, \text{SNR}) = \mathbf{P}(\varepsilon) \doteq \text{SNR}^{-d(r)},$$

with $d(r) = d_i(r) = (\rho M - r)(m - r)$.

The split NVD parallel codes in Figure 3 achieve therefore the optimal DMT of $(\rho M - r)(m - r)$.

VI. NUMERICAL RESULTS

In order to compare the performance of the split NVD parallel code with the classical NVD parallel code, we consider the case of 2 parallel 2×2 MIMO channel, *i.e.* a block fading channel with a total number of blocks equal to 2.

The structure of the NVD parallel code for this configuration is given in [8], such that

$$\mathbf{X} = \begin{pmatrix} \mathbf{\Xi} & 0 \\ 0 & \tau(\mathbf{\Xi}) \end{pmatrix} \quad (15)$$

where $\mathbf{\Xi}$ is given in (16) with $\theta = \frac{1+\sqrt{5}}{2}$, $\bar{\theta} = \frac{1-\sqrt{5}}{2}$, $\alpha = 1 + i - i\theta$, $\bar{\alpha} = 1 + i - i\bar{\theta}$ and $\zeta_8 = e^{\frac{i\pi}{4}}$. The channel matrix $\tau(\mathbf{\Xi})$ can be deduced from $\mathbf{\Xi}$ by replacing ζ_8 by $-\zeta_8$.

$$\mathbf{\Xi} = \frac{1}{\sqrt{5}} \begin{pmatrix} \alpha(s_1 + s_2\zeta_8 + s_3\theta + s_4\zeta_8\theta) & \alpha(s_5 + s_6\zeta_8 + s_7\theta + s_8\zeta_8\theta) \\ \zeta_8\bar{\alpha}(s_5 + s_6\zeta_8 + s_7\bar{\theta} + s_8\zeta_8\bar{\theta}) & \bar{\alpha}(s_1 + s_2\zeta_8 + s_3\bar{\theta} + s_4\zeta_8\bar{\theta}) \end{pmatrix}. \quad (16)$$

For the same channel model, the structure of the split NVD parallel code is such that,

$$\mathbf{X} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{\Xi}_1 & \mathbf{\Xi}_2 \\ \tau(\mathbf{\Xi}_2) & \tau(\mathbf{\Xi}_1) \end{pmatrix} \quad (17)$$

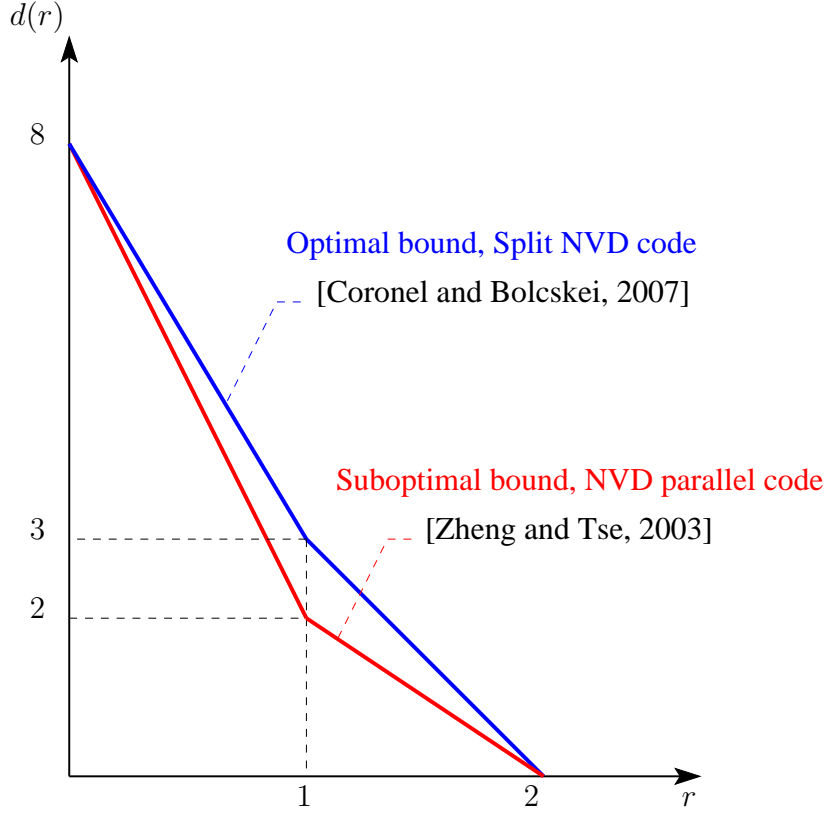


Fig. 4. The optimal DMT achievable by the NVD parallel code for the 2×2 block fading channel with $N = 2$ is $d(r) = 2(2 - r)(2 - r)$. The split code achieves the optimal DMT of the block fading channel $d(r) = (4 - r)(2 - r)$.

As we showed in previous section, the optimal DMT achievable by the NVD parallel is only $2(2 - r)(2 - r)$. However, the optimal DMT achievable by the split code is $(4 - r)(2 - r)$. These two DMT are depicted in Figure 4.

For a rate per channel use equal to 4 bpcu (resp. 8 bpcu), the symbols s_1, s_2, \dots, s_8 should be carved from a BPSK (resp. QPSK) constellation for the scheme with split code and from a QPSK (resp. 16QAM) constellation for the scheme with NVD parallel code. One should expect here that the gain provided by the use of a smaller size of constellation used in the split NVD parallel code to be compensated by the normalization factor $1/\sqrt{2}$. Due to the gain in DMT, this is not the case and the comparison of both schemes is in Figure 5. It can be easily shown there the gain of the split codes compared to the NVD parallel case is significant when the spectral efficiency of the code increases. For a small rate of 4 bpcu, a small gain can be observed. However, for the rate of 8 bpcu, approximately 5 dB of gain can be observed.

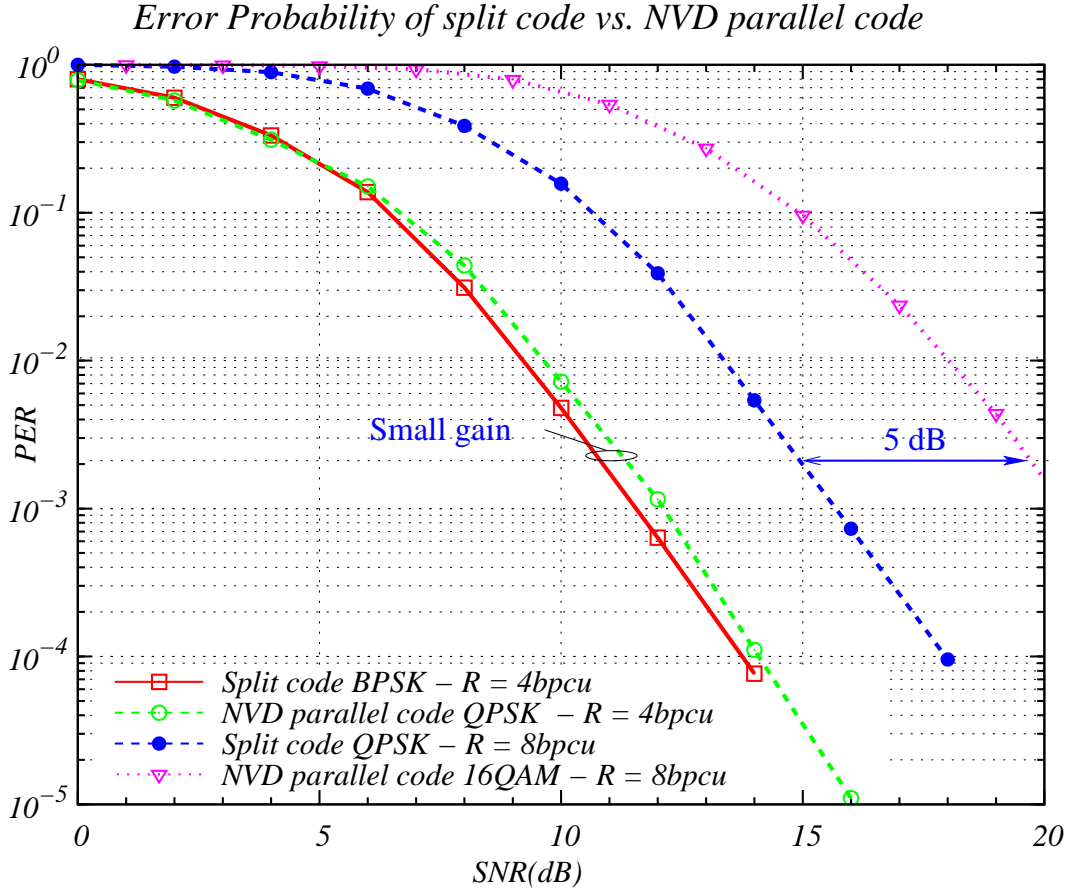


Fig. 5. Comparison of split NVD code versus NVD parallel code for a block fading MIMO channel with $N = 2$ blocks and $n_t = n_r = 2$.

VII. CONCLUSION

In this paper, we considered the class of selective fading MIMO channel where the channel is selective either in time and in frequency. Motivated by the open literature debate on the optimal achievable DMT for the block fading channel and using completely different arguments than [5] and [4], we proved here that the optimal DMT expression in [5] is achievable for all the classes of selective fading channels, including the block fading channel. Using the geometrical argument, we showed that the outage bound in [1] is not limiting for the outage probability as claimed in [6]. Moreover, a new family of split NVD parallel codes to achieve the optimal DMT in [5] for the case of time or frequency selective channels is proposed.

APPENDIX A

PROOF OF THEOREM 1

Let \mathbf{X} be the transmitted codeword, $\hat{\mathbf{X}}$ the nearest decoded codeword and $\Delta\mathbf{X}_n = \mathbf{X}_n - \hat{\mathbf{X}}_n$ the difference codeword matrix. The pairwise error probability of the correlated parallel channels is upper-bounded as following,

$$\begin{aligned} \text{PEP} &\leq \mathbb{E}_{\mathbf{H}} \exp \left(-\frac{\text{SNR}}{4n_t} \sum_{n=0}^{N-1} \|\mathbf{H}_n \Delta\mathbf{X}_n\|_{\text{F}}^2 \right), \\ &\leq \mathbb{E}_{\mathbf{H}} \exp \left(-\frac{\text{SNR}}{4n_t} \text{Tr}(\mathbf{H}_w \boldsymbol{\Theta} \mathbf{H}_w^\dagger) \right), \end{aligned} \quad (18)$$

where \mathbf{H}_w denotes the $n_r \times Nn_t$ i.i.d. $\mathcal{CN}(0, 1)$ matrix, and

$$\boldsymbol{\Theta} = (\mathbf{R}_{\mathbb{H}}^{1/2} \otimes \mathbf{I}_{n_t}) \text{diag} \left\{ \Delta\mathbf{X}_n \Delta\mathbf{X}_n^\dagger \right\}_{n=0}^{N-1} (\mathbf{R}_{\mathbb{H}}^{1/2} \otimes \mathbf{I}_{n_t})$$

is the effective codeword matrix.

Assuming that $\mathcal{X}_p(\text{SNR})$ satisfies the NVD criteria, then $\mathbf{D} = \text{diag} \left\{ \Delta\mathbf{X}_n \right\}_{n=0}^{N-1}$ is a full rank matrix with rank equals to Nn_t . The rank and the eigenvalues of the effective codeword matrix $\boldsymbol{\Theta}$ can be computed using the following lemma 2.

Lemma 2: Let \mathbf{A} be a $p \times p$ Hermitian matrix given by,

$$\mathbf{A} = \mathbf{B}(\mathbf{C}\mathbf{C}^\dagger)\mathbf{B}^\dagger,$$

where \mathbf{B} is $p \times p$ matrix with rank s , \mathbf{C} is full rank $p \times p$ matrix. Then, the matrix \mathbf{A} has the following properties:

- a) The rank of \mathbf{A} is equal to s , the rank of \mathbf{B} .
- b) The non zero eigenvalues $\lambda_k(\mathbf{A})$ of \mathbf{A} are lower bounded by,

$$\lambda_k(\mathbf{A}) \geq \lambda_1(\mathbf{B}\mathbf{B}^\dagger)\lambda_k(\mathbf{C}\mathbf{C}^\dagger). \quad (19)$$

Proof: The proof of this lemma uses the same matricial tools as [4], and is detailed in Appendix B. ■

By applying Lemma 2-a to $\boldsymbol{\Theta}$, it follows that,

$$\begin{aligned} \text{rank}\{\boldsymbol{\Theta}\} &= \text{rank}\{\mathbf{R}_{\mathbb{H}}^{1/2} \otimes \mathbf{I}_{n_t}\} \\ &= \text{rank}\{\mathbf{R}_{\mathbb{H}}^{1/2}\} \text{rank}\{\mathbf{I}_{n_t}\} = \rho n_t. \end{aligned}$$

By noticing that $\boldsymbol{\Theta}$ is not full rank, the Frobenius norm in (18) has the same distribution as $\text{Tr}\{\bar{\mathbf{H}}_w \bar{\mathbf{A}} \bar{\mathbf{H}}_w^\dagger\}$ where $\bar{\mathbf{H}}_w$ is the $n_r \times \rho n_t$ effective channel with i.i.d. entries $\sim \mathcal{CN}(0, 1)$

and $\bar{\mathbf{A}}$ is the $\rho n_t \times \rho n_t$ diagonal matrix containing the non-zero eigenvalues of the effective codeword $\mathbf{\Theta}$ bounded using Lemma 2-b such that

$$\lambda_i(\mathbf{\Theta}) \geq \sigma_{\mathbb{H}}^2 \lambda_i(\mathbf{D}\mathbf{D}^\dagger), \quad i = 1 \dots \rho n_t,$$

where $\sigma_{\mathbb{H}}^2$ is the smallest eigenvalue of $\mathbf{R}_{\mathbb{H}}$.

By following the same footsteps as in [(105) and (108) in [4]], this Frobenius norm can be bounded such that,

$$\begin{aligned} \text{Tr}\{\bar{\mathbf{H}}_w \bar{\mathbf{A}} \bar{\mathbf{H}}_w^\dagger\} &\geq \sum_{i=1}^m \lambda_i(\mathbf{H}_w \mathbf{H}_w^\dagger) \lambda_{m-i+1}(\mathbf{\Theta}) \\ &\geq \sigma_{\mathbb{H}}^2 \sum_{i=1}^m \lambda_i(\mathbf{H}_w \mathbf{H}_w^\dagger) \lambda_{m-i+1}(\mathbf{D}\mathbf{D}^\dagger) \end{aligned}$$

where \mathbf{H}_w denotes the $m \times \rho M$ Jensen channel with i.i.d. $\mathcal{CN}(0, 1)$ entries such that,

$$\mathbf{H}_w = \begin{cases} [\mathbf{H}_{w,0} \dots \mathbf{H}_{w,\rho-1}], & \text{if } n_r \leq n_t, \\ [\mathbf{H}_{w,0}^\dagger \dots \mathbf{H}_{w,\rho-1}^\dagger], & \text{if } n_r > n_t. \end{cases} \quad (20)$$

The rest of the proof uses the same technique as presented in [5], [4]. It can be deduced that if the code satisfies the NVD criteria in (9), then the error region event $E_\alpha(r, \text{SNR})$ for a given channel realisation α matches with the outage region $\mathcal{O}_\alpha^{[m, \rho M]}(r, \text{SNR})$ of the equivalent $m \times \rho M$ MIMO channel,

$$\begin{aligned} E_\alpha(r, \text{SNR}) &= \left\{ \sum_{i=1}^k \alpha_i \geq k - r, \quad k = 1, \dots, m \right\}, \\ &= \mathcal{O}_\alpha^{[m, \rho M]}(r, \text{SNR}), \end{aligned} \quad (21)$$

with α being the vector containing the eigen exponents of the channel $\mathbf{H}_w \mathbf{H}_w^\dagger$, such that $\lambda_i(\mathbf{H}_w \mathbf{H}_w^\dagger) \doteq \text{SNR}^{-\alpha_i}$.

APPENDIX B

PROOF OF LEMMA 2

As \mathbf{A} is an Hermitian matrix, its rank is equal to the rank of $\mathbf{B}\mathbf{C}$. It can be easily checked from the product matrix rank property in (22), ($\mathbf{D} \in \mathbb{C}^{a \times b}$, $\mathbf{E} \in \mathbb{C}^{b \times c}$),

$$\begin{aligned} \text{rank}\{\mathbf{D}\} + \text{rank}\{\mathbf{E}\} - b &\leq \text{rank}\{\mathbf{D}\mathbf{E}\} \\ &\leq \min \{ \text{rank}\{\mathbf{D}\}, \text{rank}\{\mathbf{E}\} \}, \end{aligned} \quad (22)$$

and the fact that \mathbf{C} is a full rank matrix, that,

$$\text{rank}\{\mathbf{B}\} + p - p \leq \text{rank}\{\mathbf{A}\} \leq \text{rank}\{\mathbf{B}\},$$

which implies that, $\text{rank}\{\mathbf{A}\} = \text{rank}\{\mathbf{B}\}$.

Using the fact that for a square matrix $\mathbf{M} \in \mathbb{C}^{a \times a}$, $\lambda(\mathbf{M}\mathbf{M}^\dagger) = \lambda(\mathbf{M}^\dagger\mathbf{M})$, implies that

$$\lambda_k(\mathbf{A}) = \lambda_k(\mathbf{C}^\dagger \mathbf{B}^\dagger \mathbf{B} \mathbf{C}).$$

Let $\mathbf{B}^\dagger \mathbf{B} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\dagger$ be the eigenvalue decomposition of $\mathbf{B}^\dagger \mathbf{B}$, with $\mathbf{\Lambda} = [\tilde{\mathbf{\Lambda}} \mathbf{0}_{p-s}]$. Then,

$$\begin{aligned} \lambda_k(\mathbf{A}) &= \lambda_k(\mathbf{C}^\dagger \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\dagger \mathbf{C}), \\ &= \lambda_k(\mathbf{\Lambda}^{1/2} \mathbf{U}^\dagger \mathbf{C} \mathbf{C}^\dagger \mathbf{U} \mathbf{\Lambda}^{1/2}). \end{aligned}$$

Let $\mathbf{\Omega} = \mathbf{U}^\dagger (\mathbf{C} \mathbf{C}^\dagger) \mathbf{U}$ and $\tilde{\mathbf{\Omega}}$ be the $s \times s$ principal submatrix of $\mathbf{\Omega}$. Then,

$$\lambda_k(\mathbf{A}) = \lambda_k(\mathbf{\Lambda}^{1/2} \mathbf{\Omega} \mathbf{\Lambda}^{1/2}), \quad (23a)$$

$$= \lambda_k(\tilde{\mathbf{\Lambda}}^{1/2} \tilde{\mathbf{\Omega}} \tilde{\mathbf{\Lambda}}^{1/2}), \quad (23b)$$

As $\tilde{\mathbf{\Lambda}}^{1/2}$ in (23b) is non singular matrix and $\tilde{\mathbf{\Omega}}$ is Hermitian, The Ostrowski theorem in [10] can be applied,

$$\lambda_k(\mathbf{A}) \geq \lambda_1(\tilde{\mathbf{\Lambda}}) \lambda_k(\tilde{\mathbf{\Omega}}), \quad (23c)$$

$$\geq \lambda_1(\mathbf{B} \mathbf{B}^\dagger) \lambda_k(\mathbf{\Omega}), \quad (23d)$$

$$= \lambda_1(\mathbf{B} \mathbf{B}^\dagger) \lambda_k(\mathbf{C} \mathbf{C}^\dagger). \quad (23e)$$

As $\tilde{\mathbf{\Omega}}$ is a $s \times s$ submatrix of the Hermitian matrix $\mathbf{\Omega}$, (23d) follows from the application of theorem 4.3.15 in [10]. Finally, (23e) follows from the fact that \mathbf{U} is unitary matrix, and therefore $\lambda_k(\mathbf{\Omega}) = \lambda_k(\mathbf{C} \mathbf{C}^\dagger)$.

REFERENCES

- [1] L. Zheng and D. Tse, "Diversity and multiplexing : A fundamental tradeoff in multiple antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073 – 1096, May 2003.
- [2] J. C. Belfiore, G. Rekaya, and E. Viterbo, "The Golden Code: A 2×2 full rate space time code with non vanishing determinants," *IEEE Transactions on information theory*, vol. 51, no. 2, pp. 1432 – 1436, April 2005.
- [3] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H.-F. Lu, "Explicit space time codes achieving the diversity multiplexing gain tradeoff," *IEEE Transactions on information theory*, vol. 52, no. 9, pp. 3869 – 3884, September 2006.
- [4] P. Coronel and H. Bölcskei, "Optimal diversity multiplexing tradeoff in selective fading MIMO channels," submitted to *IEEE Transactions on Information Theory*, 2009. Available on arxiv <http://arxiv.org/abs/0907.2391>.

- [5] —, “Diversity multiplexing tradeoff in selective fading MIMO channels,” in *IEEE International Symposium on Information Theory, Nice, France*, June 2007, pp. 2841 – 2845.
- [6] F. Lu, J. Lahtonen, R. Vehkalahti, and C. Hollanti, “Remarks on the criteria of constructing MIMO-MAC DMT optimal codes,” in *Information Theory Workshop, Cairo*, January 2010, pp. 113 – 117.
- [7] G. Durisi, U. G. Schuster, H. BÄölcskei, and S. S. (Shitz), “Noncoherent capacity of underspread fading channels,” *submitted to IEEE Transactions on information theory*, April 2008.
- [8] S. Yang, J.-C. Belfiore, and G. R. Ben-Othman, “Perfect space-time block codes for parallel MIMO channels,” in *IEEE International Symposium on Information Theory, Seattle, USA*, July 2006, pp. 1949 – 1953.
- [9] F. Lu, “Constructions of multi-block space-time codes that achieve the diversity-multiplexing tradeoff,” *IEEE Transaction on Information Theory*, vol. 54, no. 8, pp. 3790 – 3795, August 2008.
- [10] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1999.